DOCUMENT RESUME

ED 291 768 TM 011 053

AUTHOR Stufflebeam, Daniel L.
TITLE Standards of Practice for Evaluators.
PUB DATE Apr 86
NOTE 45p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, April 16-20, 1986).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Codes of Ethics; *Educational Assessment; Evaluation Utilization; Personnel Evaluation; Professional Personnel; Reliability; *Standards; Validity
IDENTIFIERS *Joint Committee on Standards for Educ Evaluation; *Standards for Evaluation Educ Prog Proj Materials

ABSTRACT

The state of the "Standards for Evaluation of Educational Programs, Projects, and Materials" is discussed. The standards were designed by a 17-member Joint Committee on Standards for Educational Evaluation to insure an ethical approach to the evaluation of educational programs and personnel. There are four major categories of standards. Utility standards are intended to guide evaluations so that they will be informative, timely, and influential. Feasibility standards recognize that an evaluation usually must be conducted in natural conditions and require that no more materials and personnel time than necessary be consumed. Propriety standards reflect the fact that evaluations affect many people in different ways; these standards are aimed at insuring that the rights of persons affected by an evaluation will be protected. Accuracy standards include those standards that determine whether an evaluation has produced sound information; these standards require that the obtained information be technically adequate and that conclusions be linked logically to the data. Each of these four categories were broken down into topic areas, from which 30 standards were derived. Once the standards are established, the evaluator must face issues associated with tradeoffs among standards, determination of the validity of standards, development of standards for personnel, and international considerations. (TJH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED291768

STANDARDS OF PRACTICE FOR EVALUATORS

by

Daniel L. Stufflebeam
Evaluation Center
Western Michigan University

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Daniel Stufflebeam
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Presented at the Annual Meeting
of the American Educational Research Association
Symposium on Ethical Issues in Evaluative Research
San Francisco, California
April 1986

TM 011 053

Professional educators, throughout the world, must evaluate their work in order to (1) obtain direction for improving it and (2) document their effectiveness. They must evaluate the performance of students, programs, personnel, and institutions. Within various countries, such evaluations have occurred at many levels: classroom, school, school district, state or province, and national system. And there have been international comparisons of the quality of education as well. The evaluations have varied enormously: in the objects assessed, the questions addressed, the methods used, the audiences served, the funds expended, the values invoked, and, to the point of this paper, their merit and worth.

In evaluations, as in any professional endeavor, many things can and often do go wrong. They are subject to bias, misinterpretation, and misapplication. They might be motivated and conducted unethically, they might address the wrong questions and/or provide erroneous information, or they might do nothing more than waste time and resources. Indeed, there have been strong charges that evaluations, in general, have failed to render worthy services (Guba, 1969; Cronbach, et al., 1980), and often, findings from individual studies have been disputed (e.g., the "Coleman, 1966 Equal Opportunity Study"). Also, evaluators have sometimes been charged with unethical practices. Clearly, evaluation itself is subject to evaluation and to efforts to assure that it serves its clients well through practice that is both technically sound and above reproach.

During the past thirty years, there have been substantial efforts in the United States to assure and control the soundness of educational evaluation services. In addition to creating professional evaluation societies and

developing preparation programs and a substantial professional literature, there have been concerted efforts to develop and apply professional standards for educational evaluation.

In the middle 1950s, the American Psychological Association joined with the American Educational Research Association and the National Council on Measurements Used in Education to develop standards for educational and psychological tests (APA, 1954; AERA/NCME, 1955); updated versions of the "Test Standards" have been published by APA in 1966, 1974, and 1985, and they have been widely used--in the courts as well as professional settings--to evaluate tests and the uses of test scores. In 1981, the Joint Committee on Standards for Educational Evaluation, whose 17 members were appointed by 12 professional societies, issued the Standards for Evaluations of Educational Programs, Projects, and Materials (which originally was commissioned to serve as a companion volume to the "Test Standards"); in 1982, the Evaluation Research Society (Rossi, 1982) issued a parallel set of program evaluation standards (intended to deal with program evaluations both outside and inside education). Currently, the Joint Committee on Standards for Educational Evaluation is developing standards for evaluations of educational personnel (which will be a companion volume to their program evaluation standards).

The different sets of standards are noteworthy because they provide: (1) operational definitions of student evaluation and program evaluation (soon to include personnel evaluation), (2) evidence about the extent of agreement concerning the meaning and appropriate methods of educational evaluation, (3) general principles for dealing with a variety of evaluation problems, (4) practical guidelines for planning evaluations, (5) widely accepted criteria for judging evaluation plans and reports, (6) conceptual frameworks by which to study evaluation, (7) evidence of progress, in the United States, toward

professionalizing evaluation, (8) content for evaluation training, and (9) a basis for synthesizing an overall view of the different types of evaluation.

Many evaluators, psychologists, and others concerned with the evaluation of education likely are aware of the "Test Standards," but might not know about the program evaluation standards or the personnel evaluation standards, which are still under development. The purpose of this paper is to provide up-to-date information about the Standards for Evaluations of Educational Programs, Projects, and Materials (hereafter called the "Program Evaluation Standards") and the more recent work of its authors, the Joint Committee on Standards for Educational Evaluation, toward developing educational personnel evaluation standards. The paper also examines the relevance of the standards for assuring that evaluation practices are ethical as well as accurate, practical, and useful.

The paper is divided into three parts: (1) an introduction to the Joint Committee's "Program Evaluation Standards", (2) an overview of the Committee's project to develop "Educational Personnel Evaluation Standards," and (3) a discussion of the relevance of the standards for addressing ethical issues in evaluation.

INTRODUCTION TO THE PROGRAM EVALUATION STANDARDS

In general, the Joint Committee devised 30 standards that pertain to four attributes of an evaluation: Utility, Feasibility, Propriety, and Accuracy. The **Utility** standards reflect a general consensus that emerged in the educational evaluation literature during the late 1960s requiring program evaluations to respond to the information needs of their clients, and not merely to address the interests of the evaluators. The **Feasibility** standards are consistent with the growing realization that evaluation procedures must be

cost-effective and workable in real-world, politically charged settings; in a sense, these standards are a countermeasure to the penchant for applying the procedures of laboratory research to real-world settings regardless of the fit. The **Propriety** standards--particularly American--reflect ethical issues, constitutional concerns, and litigation concerning such matters as rights of human subjects, freedom of information, contracting, and conflict of interest. The **Accuracy** standards build on those that have long been accepted for judging the technical merit of information, especially validity, reliability, and objectivity. Overall, then, the "Program Evaluation Standards" promote evaluations that are useful, feasible, ethical, and technically sound--ones that will contribute significantly and appropriately to the betterment of education.

Key Definitions

The "Program Evaluation Standards" reflect certain definitions of key concepts. Evaluation means the systematic investigation of the worth or merit or some object. The object of an evaluation is what one is examining (or studying) in an evaluation: a program, a project, instructional materials, personnel qualifications and performance, or student needs and performance. Standards are principles commonly accepted for determining the value or the quality of an evaluation.

Development of the Program Evaluation Standards

To ensure that the "Program Evaluation Standards" would reflect the best current knowledge and practice, the Joint Committee sought contributions from many sources. They collected and reviewed a wide range of literature. They devised a list of possible topics for standards, lists of guidelines and pitfalls thought to be associated with each standard, and illustrative cases showing an application of each standard. They engaged a group of 30 experts

independently to expand the topics and write alternative versions for each standard. With the help of consultants, the Committee rated the alternative standards, devised their preferred set, and compiled the first draft of the "Program Evaluation Standards." They then had their first draft criticized by a nationwide panel of 50 experts who were nominated by the 12 sponsoring organizations. Based upon those critiques, the Committee debated the identified issues and prepared a version which was subjected to national hearings and field tests. The results of this five-year period of development and assessments led, in 1981, to the published version of the "Program Evaluation Standards." Presently, that version is being applied and reviewed, and the Joint Committee is collecting feedback for use in preparing the next edition.

Developers of the Program Evaluation Standards

An important feature of the standards-setting process is the breadth of perspectives that have been represented in their development. The 12 organizations that originally sponsored the Joint Committee included the perspectives of the consumers as well as those who conduct program evaluations. The groups represented on the Joint Committee and among the approximately 200 other persons who contributed include, among others, those of statistician and administrator; psychologist and teacher; researcher and counselor; psychometrician and curriculum developer, and evaluator and school board member. There is perhaps no feature about the Joint Committee that is as important as its representative nature, since by definition a standard is a widely shared principle.

Format

The depth to which the Joint Committee developed each standard is apparent in the format common to all of the standards. This format starts with a descriptor--for instance, "Formal Obligation." The descriptor is followed by

a statement of the standard, e.g., "Obligations of the formal parties to an evaluation (what is to be done, how, by whom, when) should be agreed to in writing, so that these parties are obligated to adhere to all conditions of the agreement or formally to renegotiate it," and an overview, that includes a rationale for the standard and definitions of its key terms. Also included, for each standard, are lists of pertinent guidelines, pitfalls, and caveats. The guidelines are procedures that often would prove useful in meeting the standard; the pitfalls are common mistakes to be avoided; and the caveats are warnings about being overzealous in applying the given standards, lest such effort detract from meeting other standards. The presentation of each standard is concluded with an illustration of how it might be applied in an educational evaluation. The illustration includes a situation in which the standard is violated, and a discussion of corrective actions that would result in better adherence to the standard. Usually, the illustrations are based on real situations,and they encompass a wide range of different types of evaluations: e.g., small and large, formative and summative, and internal and external. One easy step to extending the applicability of the "Program Evaluation Standards" to evaluations in fields outside education would be to develop new illustrative cases drawn directly from experiences in evaluating programs outside education. Such a step might also be useful in efforts to adapt the "Program Evaluation Standards" for use in countries outside the United States.

Content of the Standards

Utility Standards. In general, the Utility Standards are intended to guide evaluations so that they will be informative, timely, and influential. These standards require evaluators to acquaint themselves with their audiences, earn their confidence, ascertain the audiences' information needs, gear evaluations

to respond to these needs, and report the relevant information clearly and when it is needed. The topics of the standards included in this category are Audience Identification, Evaluator Credibility, Information Scope and Selection, Valuational Interpretation, Report Clarity, Report Dissemination, Report Timeliness, and Evaluation Impact. Overall, the standards of Utility are concerned with whether an evaluation serves the practical information needs of a given audience.

Feasibility Standards. The Feasibility Standards recognize that an evaluation usually must be conducted in a "natural," as opposed to a "laboratory," setting, and require that no more materials and personnel time than necessary be consumed. The three topics of the Feasibility Standards are Practical Procedures, Political Viability, and Cost Effectiveness. Overall, the Feasibility Standards call for evaluations to be realistic, prudent, diplomatic, and frugal.

Propriety Standards. The Propriety Standards reflect the fact that evaluations affect many people in different ways. These standards are aimed at ensuring that the rights of persons affected by an evaluation will be protected. The topics covered by the Propriety Standards are Formal Obligation, Conflict of Interest, Full and Frank Disclosure, Public's Right to Know, Rights of Human Subjects, Human Interactions, Balanced Reporting, and Fiscal Responsibility. These standards require that those conducting evaluations learn about and abide by laws concerning such matters as privacy, freedom of information, and protection of human subjects. The standards charge those who conduct evaluations to respect the rights of others and to live up to the highest principles and ideals of their professional reference groups. Taken as a group, the propriety Standards require that evaluations be conducted

legally, ethically, and with due regard for the welfare of those involved in the evaluation, as well as those affected by the results.

Accuracy Standards. Accuracy, the fourth group, includes those standards that determine whether an evaluation has produced sound information. These standards require that the obtained information be technically adequate and that conclusions be linked logically to the data. The topics developed in this group are Object Identification, Context Analysis, Defensible Information Sources, Described Purposes and Procedures, Valid Measurement, Reliable Measurement, Systematic Data Control, Analysis of Quantitative Information, Analysis of Qualitative Information, Justified Conclusions, and Objective Reporting. The overall rating of an evaluation against the Accuracy Standards gives a good idea of the evaluation's overall truth value.

The 30 standards are summarized in Table 1.

TABLE 1 ABOUT HERE

Eclectic Orientation

The "Program Evaluation Standards" do not exclusively endorse any one approach to evaluation. Instead, the Joint Committee has written standards that encourage the sound use of a variety of evaluation methods. These include surveys, observations, document reviews, jury trials for projects, case studies, advocacy teams to generate and assess competing plans, adversary and advocacy teams to expose the strengths and weaknesses of projects, testing programs, simulation studies, time-series studies, check lists, goal-free evaluations, secondary data analysis, and quasi-experimental design. In essence, evaluators are advised to use whatever methods are best suited for

TABLE 1

SUMMARY OF THE STANDARDS

A UTILITY STANDARDS

THE UTILITY STANDARDS ARE INTENDED TO ENSURE THAT AN EVALUATION WILL SERVE THE PRACTICAL INFORMATION NEEDS OF GIVEN AUDIENCES. THESE STANDARDS ARE:

A1 Audience Identification

Audiences involved in or affected by the evaluation should be identified, so that their needs can be addressed.

A2 Evaluator Credibility

The persons conducting the evaluation should be both trustworthy and competent to perform the evaluation, so that their findings achieve maximum credibility and acceptance.

A3 Information Scope and Selection

Information collected should be of such scope and selected in such ways as to address pertinent questions about the object of the evaluation and be responsive to the needs and interests of specified audiences.

A4 Valuational Interpretation

The perspectives, procedures, and rationale used to interpret the findings should be carefully described, so that the bases for value judgments are clear.

A5 Report Clarity

The evaluation report should describe the object being evaluated and its context, and the purposes, procedures, and findings of the evaluation, so that the audiences will readily understand what was done, why it was done, what information was obtained, what conclusions were drawn, and what recommendations were made.

A6 Report Dissemination

Evaluation findings should be disseminated to clients and other right-to-know audiences, so that they can assess and use the findings.

A7 Report Timeliness

Release of reports should be timely, so that audiences can best use the reported information.

TABLE 1 (continued)

A8 Evaluation Impact

Evaluations should be planned and conducted in ways that encourage follow-through by members of the audiences.

B FEASIBILITY STANDARDS

THE FEASIBILITY STANDARDS ARE INTENDED TO ENSURE THAT AN EVALUATION WILL BE REALISTIC, PRUDENT, DIPLOMATIC, AND FRUGAL.
THESE STANDARDS ARE:

B1 Practical Procedures

The evaluation procedures should be practical, so that disruption is kept to a minimum, and that needed information can be obtained.

B2 Political Viability

The evaluation should be planned and conducted with anticipation of the different positions of various interest groups, so that their cooperation may be obtained, and so that possible attempts by any of these groups to curtail evaluation operations or to bias or misapply the results can be averted or counteracted.

B3 Cost Effectiveness

The evaluation should produce information of sufficient value to justify the resources expended.

C PROPRIETY STANDARDS

THE PROPRIETY STANDARDS ARE INTENDED TO ENSURE THAT AN EVALUATION WILL BE CONDUCTED LEGALLY, ETHICALLY, AND WITH DUE REGARD FOR THE WELFARE OF THOSE INVOLVED IN THE EVALUATION, AS WELL AS THOSE AFFECTED BY ITS RESULTS.
THESE STANDARDS ARE:

C1 Formal Obligation

Obligations of the formal parties to an evaluation (what is to be done, how, by whom, when) should be agreed to in writing, so that these parties are obligated to adhere to all conditions of the agreement or formally to renegotiate it.

C2 Conflict of Interest

Conflict of interest, frequently unavoidable, should be dealt with openly and honestly, so that it does not compromise the evaluation processes and results.

TABLE 1 (continued)

C3 Full and Frank Disclosure

Oral and written evaluation reports should be open, direct, and honest in their disclosure of pertinent findings, including the limitations of the evaluation.

C4 Public's Right to Know

The formal parties to an evaluation should respect and assure the public's right to know, within the limits of other related principles and statutes, such as those dealing with public safety and the right to privacy.

C5 Rights of Human Subjects

Evaluations should be designed and conducted, so that the rights and welfare of the human subjects are respected and protected.

C6 Human Interactions

Evaluators should respect human dignity and worth in their interactions with other persons associated with an evaluation.

C7 Balanced Reporting

The evaluation should be complete and fair in its presentation of strengths and weaknesses of the object under innvestigation, so that strengths can be built upon and problem areas addressed.

C8 Fiscal Responsibility

The evaluator's allocation and expenditure of resources should reflect sound accountability procedures and otherwise be prudent and ethically responsible.

D ACCURACY STANDARDS

THE ACCURACY STANDARDS ARE INTENDED TO ENSURE THAT AN EVALUATION WILL REVEAL AND CONVEY TECHNICALLY ADEQUATE INFORMATION ABOUT THE FEATURES OF THE OBJECT BEING STUDIED THAT DETERMINE ITS WORTH OR MERIT.
THESE STANDARDS ARE:

D1 Object Identification

The object of the evaluation (program, project, material) should be sufficiently examined, so that the form(s) of the object being considered in the evaluation can be clearly identified.

D2 Context Analysis

The context in which the program, project, or material exists should be examined in enough detail, so that its likely influences on the object can be identified.

TABLE 1 (continued)

D3 Described Purposes and Procedures

The purposes and procedures of the evaluation should be monitored and described in enough detail, so that they can be identified and assessed.

D4 Defensible Information Sources

The sources of information should be described in enough detail, so that the adequacy of the information can be assessed.

D5 Valid Measurement

The information-gathering instruments and procedures should be chosen or developed and then implemented in ways that will assure that the interpretation arrived at is valid for the given use.

D6 Reliable Measurement

The information-gathering instruments and procedures should be chosen or developed and then implemented in ways that will assure that the information obtained is sufficiently reliable for the intended use.

D7 Systematic Data Control

The data collected, processed, and reported in an evaluation should be reviewed and corrected, so that the results of the evaluation will not be flawed.

D8 Analysis of Quantitative Information

Quantitative information in an evaluation should be appropriately and systematically analyzed to ensure supportable interpretations.

D9 Analysis of Qualitative Information

Qualitative information in an evaluation should be appropriately and systematically analyzed to ensure supportable interpretations.

D10 Justified Conclusions

The conclusions reached in an evaluation should be explicitly justified, so that the audiences can assess them.

D11 Objective Reporting

The evaluation procedures should provide safeguards to protect the evaluation findings and reports against distortion by the personal feelings and biases of any party to the evaluation.

gathering information that is relevant to the questions posed by clients and other audiences, yet sufficient for assessing a program's effectiveness, costs, responses to societal needs, feasibility, and worth. It is desirable to employ multiple methods, qualitative as well as quantitative, and the methods should be feasible to use in the given setting.

Nature of the Evaluations to be Guided by the "Program Evaluation Standards"

The Joint Committee deliberately chose to limit the "Program Evaluation Standards" to evaluations of educational programs, projects, and materials. They chose not to deal with evaluations of educational institutions and personnel nor with evaluations outside education. They set these boundaries for reasons of feasibility and political viability of the project.

Given these constraints, the Joint Committee attempted to provide principles that apply to the full range of different types of studies that might legitimately be conducted in the name of evaluation. These include, for example, small-scale, informal studies that a school committee might employ to assist in planning and operating one or more workshops; as another example, they include large-scale, formal studies that might be conducted by a special evaluation team in order to assess and report publicly on the worth and merit of a statewide or national instructional program. Other types of evaluations to which the "Program Evaluation Standards" apply include pilot studies, needs assessments, process evaluations, outcome studies, cost/effectiveness studies, and meta analyses. In general, the Joint Committee says the "Program Evaluation Standards" are intended for use with studies that are internal and external, small and large, informal and formal, and for those that are formative (designed to improve a program while it is still being developed) and summative (designed to support conclusions about the worth or merit of an

object and to provide recommendations about whether it should be retained, revised, or eliminated).

It would be a mistake to assume that the "Program Evaluation Standards" are intended for application only to heavily funded and well-staffed evaluations. In fact, the Committee doubts whether any evaluation could simultaneously meet all of the standards. The Committee encouraged evaluators and their clients to consult the "Program Evaluation Standards" to consider systematically how their investigations can make the best use of available resources in informing and guiding practice.

The "Program Evaluation Standards" must not be viewed as an academic exercise of use only to well funded developers but as a code by which to help improve evaluation practice. This message is as applicable to those educators who must evaluate their own work as it is to those who can call on the services of evaluation specialists. For both groups, consideration of the "Program Evaluation Standards" may sometimes indicate that a proposed evaluation is not worthy of further consideration, or it may help to justify and then to guide and assess the study.

Tradeoffs Among the Standards

The preceding discussion points up a particular difficulty in applying the "Program Evaluation Standards." Inevitably, efforts to meet certain standards will detract from efforts to meet others, and tradeoff decisions will be required. For example, efforts to produce valid and reliable information and to generate "ironclad" conclusions may make it difficult to produce needed reports in time to have an impact on crucial program decisions, or the attempt to keep an evaluation within cost limits may conflict with meeting such standards as Information Scope and Selection and Report Dissemination. Such conflicts will vary across different types and sizes of studies, and

within a given study the tradeoffs will probably be different depending on the stage of the study (e.g., deciding whether to evaluate, designing the evaluation, collecting the data, reporting the results, or assessing the results of the study). Evaluators need to recognize and deal as judiciously as they can with such conflicts.

Some general advice for dealing with these tradeoff problems can be offered. At a macro level, the Joint Committee decided to present the four groups of standards in a particular order: Utility, Feasibility, Propriety, and Accuracy. The rationale for this sequence might be stated as "an evaluation not worth doing isn't worth doing well." In deciding whether to evaluate, it is therefore more important to begin with assurances that the findings, if obtained, would be useful, than to start with assurances only that the information would be technically sound. If there is no prospect for utility, then of course there is no need to work out an elegant design that would produce sound information. Given a determination that the findings from a projected study would be useful, then the evaluator and client might next consider whether it is feasible to move ahead. Are sufficient resources available to obtain and report the needed information in time for its use? Can the needed cooperation and political support be mustered? And, would the projected information gains, in the judgment of the client, be worth the required investment of time and resources? If such questions cannot be answered affirmatively, then the evaluation planning effort might best be discontinued with no further consideration of the other standards. Otherwise, the evaluator would next consider whether there is any reason that the evaluation could not be carried through within appropriate bounds of propriety. Once it is ascertained that a proposed evaluation could meet conditions of utility, feasibility, and propriety, then the evaluator and client would

carefully consider the accuracy standards. By following the sequence described above, it is believed that evaluation resources would be allocated to those studies that are worth doing and that the studies would then proceed on sound bases. However, this recommended sequence is not indicative of the relative importance of the four categories of standards; the Joint Committee concluded that, in general, all four categories are equally important in judging evaluation plans, activities, and reports.

There are also problems with tradeoffs among the individual standards. The Committee decided against assigning a priority rating to each standard because the tradeoff issues vary from study to study and within a given study at different stages. Instead, the Committee provided a Functional Table of Contents that is summarized in Table 2. This matrix summarizes the Committee's judgments about which standards are most applicable to each of a range of common evaluation tasks. The standards are identified down the side of the matrix. Across the top are ten tasks that are commonly involved in any evaluation. The check marks in the cells denote which standards should be heeded most carefully in addressing a given evaluation task. All of the standards are potentially applicable in all evaluations. However, the Functional Table of Contents allows evaluators to identify quickly those standards that are most relevant to certain tasks in given evaluations.

TABLE 2 ABOUT HERE

Attestation

To assist evaluators and their clients to record their decisions about applying given standards and their judgments about the extent to which each one was taken into account, the Committee provided a citation form (see Table

TABLE 2

Standards (Descriptors)	1. Decide Whether To Do A Study	2. Clarify and Assess Purpose	3. Ensure Political Viability	4. Contract the Study	5. Staff the Study	6. Manage the Study	7. Collect Data	8. Analyze Data	9. Report Findings	10. Apply Results
A1 Audience Identification	X	X	X	X		X			X	X
A2 Evaluator Credibility	X		X	X	X	X	X			
A3 Information Scope and Selection				X			X		X	
A4 Valuational Interpretation		X	X				X	X	X	X
A5 Report Clarity									X	
A6 Report Dissemination			X	X					X	X
A7 Report Timeliness				X					X	
A8 Evaluation Impact	X	X	X						X	X
B1 Practical Procedures			X				X	X		
B2 Political Viability	X		X	X	X	X	X			X
B3 Cost Effectiveness	X	X				X				
C1 Formal Obligation	X		X	X		X	X			X
C2 Conflict of Interest	X	X	X	X	X	X				X
C3 Full & Frank Disclosure				X					X	
C4 Public's Right to Know			X	X					X	X
C5 Rights of Human Subjects			X	X		X	X			X
C6 Human Interactions			X			X	X			
C7 Balanced Reporting							X		X	X
C8 Fiscal Responsibility			X	X		X				
D1 Object Identification	X	X		X			X	X	X	
D2 Context Analysis	X	X					X	X	X	X
D3 Described Purposes & Procedures	X	X	X	X		X	X		X	X
D4 Defensible Information Sources			X				X		X	
D5 Valid Measurement							X			
D6 Reliable Measurement							X			
D7 Systematic Data Control						X	X			
D8 Quantitative Analysis								X		
D9 Qualitative Analysis								X		
D10 Justified Conclusions								X	X	X
D11 Objective Reporting			X		X				X	

3). This form is to be completed, signed, and appended to evaluation plans and reports. Like an auditor's statement, the signed citation form should assist audiences to assess the merits of given evaluations. Of course, the completed citation form should often be backed up by more extensive documentation, especially with regard to the judgments given about the extent that each standard was taken into account. In the absence of such documentation, the completed citation form can be used as an agenda for discussions between evaluators and their audiences about the adequacy of evaluation plans or reports.

TABLE 3 ABOUT HERE

Validity of the Standards

In the short time since the "Program Evaluation Standards" were published, a considerable amount of information that bears on the validity of the standards has been presented. In general, this evidence supports the position that the "Program Evaluation Standards" are needed, have been carefully developed, have good credibility in the United States, and have been put to practical use. However, the assessments also point out some limitations and areas for improvement.

Bunda (1982), Impara (1982), Merwin (1982), and Wardrop (1982) examined the congruence between the "Program Evaluation Standards" and the principles of measurement that are embodied in the Standards for Educational and Psychological Tests (APA, 1974); they independently concluded that great consistency exists between these two sets of standards with regard to measurement. Ridings (1980) closely studied standard setting in the accounting and auditing fields and developed a check list by which to assess the Joint Committee

BEST COPY AVAILABLE

TABLE 3

Citation Form*

The *Standards for Evaluations of Educational Programs, Projects, and Materials* guided the development of this (check one):

request for evaluation plan/design/proposal
evaluation plan/design/proposal
evaluation contract
evaluation report
other

To interpret the information provided on this form, the reader needs to refer to the full text of the standards as they appear in Joint Committee on Standards for Educational Evaluation, *Standards for Evaluations of Educational Programs, Projects, and Materials*. New York: McGraw-Hill, 1980

The *Standards* were consulted and used as indicated in the table below (check as appropriate):

Descriptor		The Standard was deemed applicable and to the extent feasible was taken into account	The Standard was deemed applicable but was not taken into account	The Standard was not deemed applicable	Exception was taken to the Standard
A1	Audience Identification				
A2	Evaluator Credibility				
A3	Information Scope and Selection				
A4	Valuational Interpretation				
A5	Report Clarity				
A6	Report Dissemination				
A7	Report Timeliness				
A8	Evaluation Impact				
B1	Practical Procedures				
B2	Political Viability				
B3	Cost Effectiveness				
C1	Formal Obligation				
C2	Conflict of Interest				
C3	Full and Frank Disclosure				
C4	Public's Right to Know				
C5	Rights of Human Subjects				
C6	Human Interactions				
C7	Balanced Reporting				
C8	Fiscal Responsibility				
D1	Object Identification				
D2	Context Analysis				
D3	Described Purposes and Procedures				
D4	Defensible Information Sources				
D5	Valid Measurement				
D6	Reliable Measurement				
D7	Systematic Data Control				
D8	Analysis of Quantitative Information				
D9	Analysis of Qualitative Information				
D10	Justified Conclusions				
D11	Objective Reporting				

Name: _____ Date: _____
(typed)

_____
(signature)

Position or Title: _____

Agency: _____

Address: _____

Relation to Document: _____
(e.g., author of document, evaluation team leader, external auditor, internal auditor)

effort against key checkpoints in the more mature standard-setting programs in accounting and auditing. In general, she concluded that the Joint Committee had adequately dealt with four key issues: rationale, the standard-setting structure, content, and uses. Wildemuth (1981) issued an annotated bibliography with about five sources identified for each standard; these references help to confirm the theoretical validity of the "Program Evaluation Standards," and they provide a convenient guide to users for pursuing in-depth study of the involved principles. Linn (1981) reported the results of about 25 field trials that were conducted during the development of the "Program Evaluation Standards;" these confirmed that the "Program Evaluation Standards" were useful, but not sufficient guides, in such applications as designing evaluations, assessing evaluation proposals, judging evaluation reports, and training evaluators. Additionally, they provided direction for revising the "Program Evaluation Standards" prior to publication. Stake (1981) observed that the Joint Committee had made a strong case in favor of evaluation standards, but he urged a careful look at the case against standards. He offered analysis in this vein and questioned whether the evaluation field has matured sufficiently to warrant the development and use of standards.

A number of writers have examined the applicability of the "Program Evaluation Standards" to specialized situations. Wargo (1981) concluded that the "Program Evaluation Standards" represent a sound consensus of good evaluation practice, but he called for more specificity regarding large-scale, government-sponsored studies and for more representation from this sector on the Committee. (Ironically, Federal agencies had been invited to appoint representatives to the Joint Committee but declined due to potential conflicts of interest regarding their involvement in funding the effort.) Marcia Linn (1981) concluded that the "Program Evaluation Standards" contain sound advice

for evaluators in out-of-school learning environments, but she observed the "Program Evaluation Standards" are not suitable for dealing with tradeoffs between standards or settling disputes between and among stakeholders. While the "Program Evaluation Standards" explicitly are not intended for personnel evaluations, Carey (1979) examined the extent to which they are congruent with state evaluation policies for evaluating teachers; she concluded that only one standard (D11, Objective Reporting) was deemed inappropriate for judging teacher evaluations.

Burkett and Denson (1985) surveyed participants at a conference on evaluation in the health professions to obtain their judgments of the "Program Evaluation Standards." While the respondents generally agreed "...that the Standards represent a useful framework for designing evaluations and offer substantial potential for application to the evaluation of continuing education programs for the health professions," they also issued the following criticisms:

1. Crucial elements of certain standards lie outside the evaluator's professional area of control.
2. The Standards assume more flexibility, e.g., in the choice of methods of assessment, than sometimes may exist in institutional settings.
3. The Standards deal better with external evaluations than with internal, self-evaluations.
4. The Standards need to be made more useful by ordering them in the same sequence as an evaluation typically unfolds, providing more specific guidelines and examples, and adding bibliographic references.

Marsh, Newman,and Boyer (1981) used the "Program Evaluation Standards" to study the practice of educational evaluation in California and concluded the following: "(1) the standards were perceived as important ideals for the

orientation of the process and practice of evaluation; (2) the current practice of evaluation in California was perceived by professional evaluators as being, at most, of average quality; and (3) the practice of low quality evaluation was attributed to a combination of restriction of time, of political and bureaucratic coercions, and of incompetence of the evaluator."

Several evaluators from other countries have examined the "Program Evaluation Standards" for their applicability outside the United States. Nevo (1982) and Straton (1982), respectively from Israel and Australia, both concluded that while the "Program Evaluation Standards" embody sound advice, they assume an American situation--regarding level of effort and citizens' rights, for example--that is different from their own national contexts. Rodrigues, Hoffman, Barros, Arruda, and Santos (1982) published, in Portuguese, a summary and critique of the "Program Evaluation Standards" in the hope that their contribution would "...positively influence the quality of the evaluations conducted in Brazil, help in the training of educational evaluators, and help those who recommend evaluations to improve their value." Lewy, from Israel, concluded that the "Program Evaluation Standards" "...provide useful guidelines for evaluators in Israel as well as the USA," but raised questions about the adequacy of their theoretical rationale and criticized their lack of specificity.

Lewy, like Dockrell (1983), saw great possibilities for unhealthy collusion between evaluators and sponsors and disagreed with the position reflected in the "Program Evaluation Standards" that evaluators should communicate continuously with their clients and report interim findings. Dockrell also observed that evaluation in Scotland and other European countries is much more qualitatively oriented than is evaluation practice in the United States and that the "Program Evaluation Standards" do not and probably could not provide

much guidance for the perceptiveness and originality required of excellent qualitative research. Scheerens and van Seventer (1983) saw in the "Program Evaluation Standards" a useful contribution to the important need in the Netherlands to upgrade and professionalize evaluation practice; but, to promote utility in their country, they said the standards would need to be translated and illustrated at the national research policy level, as opposed to their present concentration on the individual evaluation project. Even so, they questioned whether such standards could be enforced in Holland, given the susceptibility of national research policy there to frequently changing political forces and priorities. Marklund (1983) concluded that the "Program Evaluation Standards" provides a "...good check list of prerequisites for a reliable and valid evaluation," but that "...due to differences in values of program outcomes, such standards do not guarantee that the result of the evaluation will be indisputable." Overall, the main value of the "Program Evaluation Standards" outside the United States appears to be as a useful reference for stimulating discussion of the need for professionalizing evaluation and the range of issues to be considered.

Six studies were conducted to examine the extent to which the "Program Evaluation Standards" are congruent with the set of program evaluation standards that was recently issued by the Evaluation Research Society. Rossi (1982), Cordray (1982), Braskamp and Mayberry (1982), Stufflebeam (1982), McKillip (1983), and Stockdill (1984) found that the two sets of standards are largely overlapping.

Overall, the literature on the "Program Evaluation Standards" indicates considerable support for these standards. They are seen to fill a need. They are judged to contain sound content. They have been shown to be applicable in a wide range of American settings. They have been applied successfully. They

are consistent with the principles in other sets of standards. And they are subject to an appropriate process of review and revision. But, by no means are they a panacea. Their utility is limited, especially outside the United States. And several issues have been raised for consideration in subsequent revision cycles.

STANDARDS FOR EVALUATIONS OF EDUCATIONAL PERSONNEL

An initial decision in developing the "Program Evaluation Standards" was to exclude the area of personnel evaluation. One reason was that developing a whole new set of standards for program evaluation presented a sufficiently large challenge; another reason was that members of the Committee believed that teachers' organizations would not support development of standards for evaluations of personnel. Also, in 1975 when the Joint Committee was formed, there was little concern for increasing or improving the evaluation of educational personnel.

The Decision to Develop Educational Personnel Evaluation Standards

In 1984, a number of factors led to the Joint Committee's decision to develop standards for evaluations of educational personnel. The Committee had successfully developed the "Program Evaluation Standards" and felt capable of tackling the personnel evaluation standards issue. They were also convinced that personnel evaluation in education was greatly in need of improvement. Moreover, they saw this need as urgent, because of the great increase in the development of systems for evaluating teachers and because of the great turmoil and litigation that accompanied the expansion of educational personnel evaluation activity. Moreover, they believed that the major teachers' organizations would support the development of professional standards that could be used to expose unsound plans and programs of personnel evaluation.

Expansion of the Joint Committee

In the course of deciding to develop the educational personnel evaluation standards, the Committee also decided to expand its membership to ensure that its members reflected relevant perspectives on evaluations of educational personnel as well as evaluations of educational programs. Additions to the Committee included representatives from the American Association of School Personnel Administrators, the American Federation of Teachers, and the American Association of Secondary School Principals, as well as individual members-at-large with expertise in litigation in personnel evaluation and research on teacher evaluation. New appointments by sponsoring organizations also included the perspectives of industrial/organizational psychology and traditionally underrepresented groups. The 18-member Committee continues to include a balance between the perspectives of educational practitioners and evaluation specialists. The membership and organizational affiliations of the Joint Committee are listed in Table 4.

TABLE 4 ABOUT HERE

Validation Panel

An independent validation panel provides further perspective and checks and balances on the work of the Committee. This group is led by Dr. Robert Linn and includes persons representing the following perspectives: law, research on teaching, personnel psychology, international education, educational research, psychometrics, philosophy, teaching, school district superintendency, and school principalship. Their charge is to monitor and evaluate

TABLE 4

MEMBERS OF THE JOINT COMMITTEE ON STANDARDS FOR EDUCATIONAL EVALUATION
MAY 1986

Chair

Daniel L. Stufflebeam (Western Michigan University)

Committee Members

James Adams (Indianapolis Public Schools), representing the American Association of School Administrators

Ralph Alexander (University of Akron), representing the American Psychological Association

Beverly Anderson (Education Commission of the States), representing the Education Commission of the States

Esther Diamond (Educational and Psychological Consultant), representing the Association for Measurement and Evaluation in Counseling and Development

A. Keith Esch (Wichita Public Schools), representing the American Association of School Personnel Administrators

Ronald K. Hambleton (University of Massachusetts), representing the National Council on Measurement in Education

Philip L. Hosford (New Mexico State University), representing the Association for Supervision and Curriculum Development

William Mays, Jr. (Michigan Elementary and Middle School Principals Association), representing the National Association of Elementary School Principals

Diana Pullin (Michigan State University), member-at-large

Marilyn Rauth (American Federation of Teachers), representing the American Federation of Teachers

James Sanders (Western Michigan University), representing the American Evaluation Association

Sheila Simmons-Merrick (National Education Association), representing the National Education Association

Robert Soar (University of Florida), member-at-large

Scott Thomson (National Association of Secondary School Principals), representing the National Association of Secondary School Principals

JoAnn Wimmer (Logan, Utah), representing the National School Boards Association

Linda Winfield (New Castle County, Delaware School District Consortium), representing the American Educational Research Association

Arthur Wise (Rand Corporation), member-at-large

the work of the Committee and ultimately to publish an independent evaluation of the standards for evaluations of educational personnel. The Validation Panel's main clients are those groups who might for a variety of reasons want independent assessments of the appropriateness, quality, and potential utility of the standards. The membership of the Validation Panel* is listed in Table 5.

TABLE 5 ABOUT HERE

The Guiding Rationale

It is appropriate for the Joint Committee to deal with personnel evaluation as well as program evaluation. Both types of evaluation are prevalent in education, and both are vitally important for assuring the quality of educational services. Practically and politically it is usually necessary to conduct these two types of evaluation separately. But logically, they are inseparable.

Practice and literature have lodged responsibility for personnel evaluation with supervisors and administrators and have created expectations that program evaluators will not evaluate the performance of individuals as such. Program evaluators might provide some technical advice for developing a sound system of personnel evaluation and might even evaluate the personnel evaluation system itself; but they have preferred, and often have insisted on,

---

*The teacher and principal positions on the Panel are currently being filled.

TABLE 5

VALIDATION PANEL
Joint Committee on Standards for Educational Evaulation

Member	Perspective
Dr. Margret Buchmann Michigan State University	Philosophy of education
Dr. Constance Clayton School District of Philadelphia	Administration
Dr. Edmund Gordon Yale University	Research on teaching
Dr. Bruce Gould U. S. Air Force	Personnel psychology
Dr. Thomas Kellaghan St. Patrick's College Dublin, Ireland	International education
Dr. Robert Linn (Chair) University of Illinois at Urbana	Educational psychology
Dr. Perry Zirkel Lehigh University	Law in education

staying out of the role of directly evaluating individual personnel. To do otherwise would stimulate fear about the power and motives of evaluators,and would undoubtedly generate much resistance on the part of principals and teachers, leading in turn to lack of cooperation in efforts to evaluate programs. Thus, program evaluators typically have avoided involvements with personnel evaluation. They have emphasized instead the constructive contributions of program evaluation, and they have promised as much anonymity and confidentiality as they could to teachers and administrators in the programs being evaluated. On the whole, efforts to separate personnel and program evaluation in school districts have remained in vogue.

But a basic problem remains: namely, it is fundamentally impossible to remove personnel evaluation from sound program evaluation. A useful program evaluation must determine whether a program shows a desirable impact on the rightful target population. If the data reveal otherwise, the assessment must discern those aspects of a program that require change to yield the desired results. Inescapably, then, program evaluators must check the adequacy of all relevant instrumental variables, including the personnel. The rights of teachers and administrators must be respected, but evaluators must also protect the rights of students to be taught well and of communities to have their schools effectively administered.

However, personnel evaluation is too important and difficult a task to be left exclusively to the program evaluators. Many personnel evaluations are conducted by supervisors who rarely conduct formal program evaluations. Also, state education departments and school districts are heavily involved, apart from their program evaluation efforts, in evaluating teachers and other educators for certification, selection, placement, promotion, tenure, merit, staff development, and termination.

Undesirably, the literatures and methodologies of program evaluation and personnel evaluation are distinct. The work of the Joint Committee in both areas affords a significant opportunity to bring a concerted effort to bear on synthesizing these fields and coordinating the efforts of program evaluators and personnel evaluators for the betterment of educational service.

The Developmental Process

To achieve its goals for developing standards for personnel evaluations, the Joint Committee is employing the approach it found successful in the development of the "Program Evaluation Standards." They have collected and studied an enormous amount of information about educational personnel evaluation and have developed a tentative set of topics for personnel evaluation standards. A panel of writers, nominated by the 14 sponsoring organizations, wrote multiple versions of each proposed standard. The Joint Committee evaluated the alternative versions and decided which aspects of each standard would be included in the initial review version of the Educational Personnel Standards book. The first draft of the book is currently being critiqued by a national review panel and an international review panel. The Joint Committee will use the critiques to develop a semifinal version of the book. That version will be field tested and subjected to hearings conducted throughout the United States. The results will be used to develop the final publication version of the Educational Personnel Evaluation Standards. Publication is expected in 1988.

Contents of the Standards

After reviewing a great deal of material on personnel evaluation, the Joint Committee decided that the four basic concerns of Utility, Feasibility, Propriety, and Accuracy are as relevant to personnel evaluation as they are to program evaluation. Some of the topics for individual standards are likewise

the same, e.g., valid measurement and reliable measurement. However, there are important differences in the two sets of topics. For example, Full and Frank Disclosure, a program evaluation standard, hasn't surfaced in the personnel evaluation standards; and Service Orientation, a key entry in the personnel evaluation standards (requiring that evaluators show concern for the rights of students to be taught well), wasn't among the "Educational Program Evaluation Standards." In general, much work remains to be done before the contents of the first edition of the Educational Personnel Evaluation Standards will be finalized.

International Involvements and Implications

The Committee desires to stay in touch with international groups that are involved in evaluations of educational personnel so that it can benefit by the experiences in other countries and share what it learns from this project with interested groups in those countries. Accordingly, an Irish psychologist serves on the Validation Panel to add an international perspective, and the Committee has engaged an International Review Panel to evaluate the first draft of the standards. The Committee will also report its progress to international audiences through a periodic newsletter. However, they believe the standards must concentrate on the relevant U.S. laws and personnel evaluation systems; consequently, the personnel evaluation standards might not transfer well to other cultures.

IMPLICATIONS OF STANDARDS FOR ADDRESSING ETHICAL ISSUES IN EVALUATIONS

As seen in the preceding sections, the standards for both programs and personnel offer some protections to various parties to an evaluation in the realm of propriety. This section discusses the need for propriety standards,

describes some of the relevant issues, and assesses the potential contributions and limitations of professional standards vis-a-vis ethical issues.

Because evaluation is a specialized field of practice, and because it is often imposed as a condition for funding or holding a job, many persons and organizations have little choice but to use (or be subjected to) its services. Consequently, they are dependent on the work of evaluation specialists and are at risk to the extent that the evaluators are incompetent, careless, or unethical.

To the point of the symposium for which this paper was written, there have been many charges and confirmed cases of unethical practices in evaluation work. These include falsifying results; maliciously defaming a person or organization; violating a person's right to privacy; accepting an assignment to advocate or attack something according to the interests of the client; covering up negative findings; overstating a criticism in order to gain national attention; or exposing subjects, without their knowledge or consent, to possible harm by their participation in a study. I have no doubt that each presenter in the symposium and many members of the audience could identify concrete examples of such abuses. Clearly, evaluators need standards of practice that deal with ethical issues as well as other issues, such as those concerned with technical adequacy, utility, and feasibility. However, incorporation of ethical considerations into professional standards for evaluators has not been easy or extensive, as the history of evaluation reveals.

Historical Perspective on Evaluation Standards and Ethics

Among the first systematic presentations of criteria for judging evaluation studies were those of internal validity and external validity, as articulated by Campbell and Stanley in 1963. Their recommendations tended to restrict criteria for judging evaluations to technical matters and thereby

drew interest away from other issues in evaluation work, especially propriety and utility. Subsequent treatments expanded the suggested criteria to include utility and efficiency as well as technical adequacy. This expansion was seen in a 1971 book by the Phi Delta Kappa Study Committee on Evaluation (Stufflebeam, et al., 1971). Their recommendations cast evaluation in more of an instrumental role than had the recommendations by Campbell and Stanley. However, the PDK group's recommendations did not address such ethical concerns as protection of human subjects, censorship of reports, and due process. Evaluators didn't write seriously about such ethical criteria for evaluations until the middle 1970s, well after relevant laws had been passed and enforced. Only then did the literature of evaluation begin to delve into concerns of human rights, freedom of information, and similar ethical problem areas.

The current status of the evaluation field in dealing with ethical concerns is partially reflected in the recent work of the Joint Committee on Standards for Educational Evaluation.

The Committee's two sets of standards books each spell out about eight propriety standards, as well as about twenty more in the areas of utility, feasibility, and accuracy. The standards provide principles of good practice, give examples of malpractice, and offer practical guidelines. They are intended to be used by evaluators to check their plans and reports against a wide range of public criteria. The standards also provide carefully developed content for training new or aspiring members of the evaluation field. In addition, they provide a tool that clients can use to write sound contracts for evaluation services and to expose and document poor or unethical practice. In general, the standards promote ethical practices in evaluation.

Some examples of the constructive recommendations in the standards related to propriety issues are as follows:

- identify all groups that are entitled to the findings and provide them with access to the reports
- search out and openly address conflicts of interest
- provide in advance for protecting the rights of those who will be affected by the study
- report both strengths and weaknesses and provide direction for improvement
- recommend actions that are in the best interests of students and other clients
- treat participants in evaluations with respect and dignity
- and, as a last example, negotiate contracts to govern the evaluation work and to help assure that the advance understandings and agreements are remembered and implemented

Although these recommendations may seem obvious, it is surprising how often evaluators get into difficulty by ignoring one or more of the recommendations. If taken seriously and applied, the standards should assist the evaluation field to gain stature as a respected and trusted profession.

Problems in Identifying and Addressing Ethical Problems in Evaluation

But, despite the progress made in defining standards of practice in evaluation, there are many unresolved ethical issues. Moreover, the evaluation profession is immature and isn't well qualified--by experience, profession-wide deliberations, or organization--to ferret out and address the full range of relevant ethical issues in evaluation work. The complexity of such issues is seen in six general concerns that are especially problematic.

First, in in evaluation work there are many naturally occurring **conflicts of interest.** The client often would be happiest with an expedient approach to goal achievement. The evaluator wants to get paid and hopes to be rehired. While many people might informally express concerns about a program or educator, those same persons often refuse to go on record with their complaints. Also, some evaluators are zealots in their support of particular methods.

Clearly, vested interests, such as those mentioned, can influence evaluations to produce biased results. The Joint Committee has tried to help reduce the bad effects of conflicts of interest by drawing attention to the issue, describing its characteristics, offering some recommendations for identifying and addressing potential conflict of interest problems, and, when needed, providing a basis for exposing conflicts of interest after the fact. But, in general, conflicts of interest are an inevitable part of the evaluation territory.

A second course of difficulty in evaluation is the great amount of **suboptimization** that occurs. This is seen especially in the lack of integration of the subfields of student evaluation, program evaluation, and personnel evaluation. A sort of ends-justifies-the-means mentality has helped to keep these subfields separate. If student test scores are judged good, then there is a tendency not to look at program or personnel, even though they may be deficient or harmful in their application. If one does evaluate a program, a frequent tendency is to grant immunity from scrutiny to the staff so as to obtain their good will and cooperation. On the other side, evaluations of personnel are frequently done in relation to a union contract and a job description, but in isolation from their roles in particular programs or their effects on students or clients. Because evaluations of personnel, programs, and students are typically done separately, evaluators often fail to address the full range of questions that might reveal improprieties. Clearly, suboptimizations help to make work manageable, but they can also help to obscure practices that are harmful or unjust. The field must find better ways to integrate evaluations of programs, personnel, and students. The development of standards in these areas provides one avenue for the pursuit of the needed integration.

A third source of difficulty in assuring ethical evaluations is an attitude among evaluators of **no harm, no foul.** According to this position, evaluators need not be concerned with professional standards if they and others don't see that clients are being harmed. Probably this type of attitude was partially responsible for evaluators' failure to address concerns about equity, due process, rights of human subjects, and censorship until after the enactment of relevant legislation and, in many cases, after the legislation was tested in the courts. Of course, waiting for government to identify and address injustices is an efficient way of identifying ethical issues that should be addressed by standards. But it is not a proper stance for professionals whose obligation is to provide the best and most ethical service possible. The Joint Committee and its standards-setting process represents one means for the evaluation field to become more proactive in identifying and addressing ethical issues.

A fourth problem area concerns **tradeoffs among standards.** It seems clear that few, if any, evaluations could simultaneously meet all the relevant standards. Therefore, evaluators and their clients must compromise between such conflicting standards as providing constructive feedback so that educators might do better, and improving learning experiences for students through such actions as helping to remove an incompetent, or otherwise harmful, administrator or teacher. Which of such conflicting standards should be given precedent, in general and in specific cases? The Joint Committee has found no easy answers to these questions. Instead, they have recommended that clients and evaluators systematically seek out and adjudicate such tradeoff problems, that they faithfully implement their decisions, and that they subject their tradeoff decisions and actions to third-party reviews. While not completely

satisfying, these answers help to emphasize that evaluation is a most complex enterprise that requires careful and audited judgments.

The fifth area of concern is that professional **standards are vulnerable to misuse.** The advice of the Joint Committee standards is mainly general. There are no specific rules for resolving the inevitable conflicts among the standards, and the standards carry no penalties for violation. In the face of these limitations, the effectiveness of the standards is largely dependent on the good intentions of evaluators and the thoughtful deliberations and wise judgments of those clients, auditors, and evaluators who apply the standards. Unfortunately, it is possible for evaluators to apply the standards superficially and to use them as a cloak to cover up their poor service or even malpractice. Like a hammer or any other tool, the standards can be misused.

The final difficulty follows from the fifth one. It is that **standards are insufficient by themselves** to ward off or treat ethical issues in evaluation. Standards are only one component of the professional initiatives that are needed to help assure that evaluation practices are ethical. Considering the experiences of more mature professions, the evaluation field needs to consider a range of special means of enforcing its principles of practice. For example, it could accredit worthy training programs and set up examinations and other mechanisms for certifying and/or licensing evaluators; such steps would aid clients to identify evaluators who are appropriately qualified. In addition, a group such as the American Evaluation Association might define sanctions for malpractice, set up a practice review board to hear charges, adopt procedures for carrying out the decisions of the review board, and subsequently use those developments to help shape up or throw out the bad actors in the evaluation field. Probably the young evaluation profession is not close to introducing such strong measures, but they might pursue such

steps in the future. In the meantime, a more realistic practice for the evaluation field is to increase its use of third-party audits or meta-evaluations, and the Joint Committee standards provide widely shared principles and recommendations to help guide such assessments.

Overall Assessment of the Standards Vis-A-Vis Ethical Issues

Professional standards provide one mechanism for promoting ethical practice in evaluation. Their greatest potential impact is on those evaluators who have a strong sense of moral responsibility and who are seeking ways to improve their services. Professional standards also provide some help to clients who want to know whether or not an evaluation proposal or report is sound, and they offer to the profession a partial basis for policing its own ranks. However, the Joint Committee standards are not, and never will be, the final word on what constitutes good and ethical evaluation service. They will always be only a dated approximation of ideals for the field, a negotiated set of general agreements. Consequently, they must be periodically reviewed and updated. Also, professional standards are not sufficient by themselves to ward off or deal with unethical practices. In addition to setting standards, the evaluation profession needs to consider measures such as certification, practice review boards, and defined sanctions. Finally, while standards and symposia at professional meetings won't resolve the ethical problems in evaluation work, they do serve to draw attention to a wide range of relevant issues. And that's one important step towards making evaluations more ethical.

CLOSING

Increasingly, evaluation is becoming a formalized field of practice. Its services are complex and costly and it has the potential to do harm as well as to promote progress. Since 1975, evaluators and educators have pursued a

concerted effort to define standards of sound practice--initially with respect to program evaluations and more recently in the area of personnel evaluations. The standards have been defined through the efforts of the Joint Committee on Standards for Educational Evaluation, whose members were appointed by fourteen professional societies. The purpose of this paper has been to present an update on the work of the Committee and particularly to discuss the relevance of their standards for addressing ethical issues in evaluation. The pervasive message is that the Joint Committee standards are an important but not sufficient means of making educational evaluations useful, feasible, accurate, and ethical.

REFERENCES

American Educational Research Association & National Council on Measurements Used in Education. (1955). Technical recommendations for achievement tests. Washington, DC: National Education Association.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1985). Standards for educational and psychological testing. Washington, DC: American Psychological Association, Inc.

American Psychological Association. (1954). Technical recommendations for psychological tests and diagnostic techniques. Washington, DC: American Psychological Association.

American Psychological Association. (1966). Standards for educational and psychological tests and manuals. Washington, DC: American Psychological Association, Inc.

American Psychological Association. (1973). Ethical principles in the conduct of research with human participants. Washington, DC: American Psychological Association, Inc.

American Psychological Association. (1974). Standards for educational and psychological tests (Rev. Ed.). Washington, DC: American Psychological Association.

Braskamp, L. A. & Mayberry, P. W. (1982). A comparison of two sets of standards. Paper presented at the joint annual meeting of the Evaluation Network and Evaluation Research Society, Baltimore, MD.

Bunda, M. (1982). Concerns and techniques in feasibility. Paper presented at the annual meeting of the National Council on Measurement in Education, New York.

Burkett, Deborah & Denson, Teri. (1985). Another view of the standards. In Abrahamson, Stephen (Ed.) Evaluation of continuing education in the health professions. Boston: Kluwer-Nijhoff Publishing.

Campbell, D. T. & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (ed.), Handbook of Research on Training. Chicago: Rand McNally.

Carey, L. (1979). State-level teacher performance evaluation policies. Inservice Centerfold. New York: National Council on State and Inservice Education.

Coleman, J. S.; Campbell, E. Q.; Hobson, C. J., et al. (1966). Equality of equal educational opportunity. Washington, DC: Office of Education, U. S. Department of Health, Education, and Welfare.

Cordray, D. (1982). An assessment of the utility of the ERS standards. In P.H. Rossi (Ed.) Standards for Evaluation Practice. New Directions for Program Evaluation, No. 15. San Francisco: Jossey-Bass.

Cronbach, L. J. (1980). Toward reform of program evaluation. San Francisco: Jossey-Bass Publishers.

Division of Industrial-Organizational Psychology, American Psychological Association. (1980). Principles for the validation and use of personnel selection procedures: Second edition. Berkeley, CA: American Psychological Association, Division of Industrial-Organizational Psychology.

Dockrell, W. B. (1983). Applicability of standards for evaluations of educational programs, projects, and materials. Presentation at the annual meeting of the American Educational Research Association, Boston.

Impara, J. C. (1982). Measurement and the utility standards. Paper presented at the meeting of the National Council for Measurement in Education, New York.

Joint Committee on Standards for Educational Evaluation. (1981). Standards for evaluations of educational programs, projects, and materials. New York: McGraw-Hill.

Lewy, Arieh. (1983). Evaluation standards: Comments from Israel. Presentation at the annual meeting of the American Educational Research Association, Boston.

Linn, Marcia. (1981). Standards for evaluating out-of-school learning. Evaluation News, 2 (2), 171-176.

Linn, R. L. (1981). A preliminary look at the applicability of the educational evaluation standards. Educational Evaluation and Policy Analysis, 3, 87-91.

McKillip, J. & Garberg, R. A further examination of the overlap between ERS and Joint Committee evaluation standards. Unpublished paper. Carbondale, IL: Southern Illinois University, Department of Pscyhology.

Margh, David D., Newman, Warren B., & Boyer, William F. (1981). Comparing ideal and real: A study of evaluation practice in California using the Joint Committee's evaluation standards. Paper presented to the annual meeting of the American Educational Research Association, Los Angeles.

Marklund, Sixten. (1983). Applicability of Standards for evaluations of educational programs, projects, and materials in an international setting. Presentation at the annual meeting of the American Educational Research Association, Boston.

Merwin, J. C. (1982). Measurement and propriety standards. Paper presented at the meeting of the National Council for Measurement in Education, New York.

Nevo, D. (1982). Applying the evaluation standards in a different social context. Paper presented at the 20th Congress of the International Association of Applied Psychology, Edinburgh, Scotland.

Ridings, J. M. (1980). Standard setting in accounting and auditing: Considerations for educational evaluation. Unpublished dissertation. Kalamazoo: Western Michigan University.

Rodrigues de Oliveira, Terezinha; Hoffman, Jussara Maria Lerch; Barros, Raimundo Faco; Arruda, Nilce Fatima Correa; & Santos, Raulo-Ruas. (1981). Standards for evaluation of educational programs, projects, and materials. Unpublished paper of the Department of Education of the Federal University of Rio de Janeiro (UFRJ).

Rossi, Peter H. (Ed.) (1982). Standards for evaluation practice. San Francisco: Jossey-Bass, Inc.

Scheerens, J. and van Seventer. (1983). Political and organizational preconditions for application of the standards for educational evaluation. Presentation at the annual meeting of the American Educational Research Association, Boston.

Stake, R. (1981). Setting standards for educational evaluators. Evaluation News 2 (2), 148-152.

Stockdill, Stacey Hueftle. (October 1984). The appropriateness of the evaluation standards for business evaluations. Presentation at the Evaluation Network/Evaluation Research Society Joint meeting. San Francisco.

Straton, R. B. (1982). Appropriateness and potential impact of programme evaluation standards in Australia. Paper presented at the 20th International Congress of Applied Psychology, Edinburgh, Scotland.

Stufflebeam, D. L., et al. (1971). Educational evaluation and decision-making. Ithaco, IL: Peacock Publishers.

Stufflebeam, D. L. (1982). An examination of the overlap between ERS and Joint Committee standards. Paper presented at the Annual Meeting of the Evaluation Network, Baltimore, MD.

Wardrop, J. C. (1982). Measurement and accuracy standards. Paper presented at the meeting of the National Council for Measurement in Education, New York.